

Position Estimation for a Mobile Robot Using Vision and Odometry

Frédéric Chenavier *

James L. Crowley **

* LETI-DSYS, CEA-CENG, 85X, F38041 Grenoble Cedex

** LIFIA-IMAG, INPG, 46 Av. Félix Viallet, F38031 Grenoble Cedex

Abstract

In this paper, we describe a method for locating a mobile robot moving in a known environment. This technique combines position estimation from odometry with observations of the environment from a mobile camera. Fixed objects in the world provide landmarks which are listed in a data base. The system calculates the angle to each landmark and then orients the camera. An extended Kalman filter is used to correct the position and orientation of the vehicle from the error between the observed and estimated angle to each landmark. Results from experiments in a real environment are presented.

1. Introduction

Location is basic to navigation. Various techniques have been described for estimating the position and orientation of a vehicle. These include odometry [4] [8], position estimation from known landmarks [6]-[11], and localization with respect to the global environment [2], [3] and [5].

The technique presented here has been developed for an intervention robot for a nuclear site. For such a robot it is not possible to equip the site with beacons. Thus we have developed a technique that uses naturally occurring structure as beacons. A class of structural elements that is both salient and view invariant is provided by vertical edges. Such edges are provided by doors, pipes (average of the two edges) and corners, and may be easily extracted by simple image processing. The position of these landmarks is stored in a data base and observations are made with a camera mounted on a steerable rotating platform.

In the classical solution to position estimation from beacons, it is necessary to observe three beacons simultaneously to obtain a non-redundant estimation of position and orientation. An extended Kalman Filter (EKF) can be used to update a position estimation from a single observation. However, the EKF requires a derivative of a non-linear function, which requires an estimation of the current position and orientation. In this paper we provide the mathematical analysis of the use of an extended Kalman filter for estimation from odometry and for estimation from observation of known vertical edges. We demonstrate experimentally that such an approach is reliable.

The technique presented in this paper builds naturally on previous results from a number of authors. For example, Chen and Tsai have shown how camera position can be recovered with an error less than 5% of the robot-landmark distance, by locating the image of a set of straight lines (trihedron, polygon, prism) [7]. Based on the location of vertical segments, Sugihara and Krotkov proposed an algorithm able to take advantage of a redundancy in measurements [10] [11]. Kriegman demonstrated position estimation from observation of vertical edges in a hall-way [9]. Techniques have been presented to use odometry and vision to locate standard patterns, with vision providing position estimation [8]. The architecture described in [6] expressed robot position in term of probability, and visual landmark detection should be used to correct it. The use of a Kalman filter to fuse odometry and ultrasonic ranging in order to update the position of a mobile robot has been described in [3] and in [5].

In much of the previous work, odometry has been used in a very ad-hoc way, while the mathematical analysis of the technique has been neglected. In addition, we continue

to encounter strong skepticism and confusion from engineers and scientists over the feasibility of the use of an EKF for position estimation from odometry and vision. In the technique described below, odometry provides a position estimation with a good short term accuracy (drifting of a few percent of displacement), while observation of the angle to known landmarks provides a correction. The contribution of this paper is

- 1) To present the mathematical analysis behind such an approach, and
- 2) To demonstrate experimentally that the technique is robust and reliable.

Robot position is represented by the random variable $X_t = (x_R, y_R, \theta_R)^T$, where x_R, y_R is the vehicle position and θ_R is the robot heading. The detection of L landmarks (one or more sets of vertical edges) provides M measurements ($M \geq L$), which can be used to correct this estimation and to reduce its associated uncertainty. For a single observation of a simple landmark, position is under-determined ($M < 3$). An extended Kalman filter permits this observation to provide a constraint on the estimated position.

The following section presents our model for the odometry uncertainty. In section 3, the measurements derived from the vision system are described. Their fusion with the odometric data through EKF are explained in section 4. Experimental results are included in section 5, and are followed by conclusions in section 6.

2. Modeling odometry

The two back wheels of our vehicle are equipped with encoders. At each sampling period, ΔT , the signed integration of the encoder pulses provides an estimate of angular displacement (respectively $\Delta\alpha_R$ and $\Delta\alpha_L$) during the period. The corresponding translation (ΔS) and rotation ($\Delta\theta$), measured with respect to the mid-point of the back axis, are given by [4] as:

$$\Delta S = R(\Delta\alpha_R + \Delta\alpha_L) / 2 \quad (1)$$

$$\Delta\theta = R(\Delta\alpha_R - \Delta\alpha_L) / B \quad (2)$$

where R is the right and left wheel radius, and B the wheel base.

These exact values ΔS and $\Delta\theta$ can be used to write a first order estimation equation for $X_t = (x_R, y_R, \theta_R)^T$, which expresses $X_{t+\Delta t}$ as a function of X_t :

$$\begin{bmatrix} x_{t+\Delta t} \\ y_{t+\Delta t} \\ \theta_{t+\Delta t} \end{bmatrix} = \begin{bmatrix} x_t + \Delta S \cos(\theta_t + \Delta\theta/2) \\ y_t + \Delta S \sin(\theta_t + \Delta\theta/2) \\ \theta_t + \Delta\theta \end{bmatrix} + W_{t,\Delta t} \quad (3)$$

This equation approximates trajectory as a sequence of constant curvature segments of length ΔS . Its precision requires that the curvature of the trajectory, given by $\frac{\Delta\theta}{\Delta S}$ be constant during the sample interval, ΔT . Such an approximation requires that the sampling period ΔT be sufficiently small with respect to the vehicle translational and rotational acceleration.

$W_{t,\Delta t}$ designates an external observation noise due to deformation of wheel radius, wheel slippage, vibrations, and other unknown errors. $W_{t,\Delta t}$ is supposed to be Gaussian, with an expectation or mean of 0 and a covariance matrix given by Q_t :

$$E[W_{t,\Delta t}] = 0 \quad (4)$$

$$E[W_{t,\Delta t} W_{t,\Delta t}^T] = Q_t \quad (5)$$

This amounts to considering $W_{t,\Delta t}$ as an integration of Gaussian noise during ΔT [1].

The system equation (3) may be rewritten as:

$$X_{t+\Delta t} = F(X_t, \Delta S, \Delta\theta) + W_{t,\Delta t} \quad (6)$$

For updating $X_{t+\Delta t}$, we only have access to estimates of the different variables (written \hat{X}_t for the estimate of X_t , etc...) so that estimated state, $\hat{X}_{t+\Delta t}$, is given by:

$$\hat{X}_{t+\Delta t} = F(\hat{X}_t, \Delta S, \Delta\theta) \quad (7)$$

The covariance matrix $C_{t+\Delta t}^X$, defined by $C_{t+\Delta t}^X = E[(X_{t+\Delta t} - \hat{X}_{t+\Delta t})(X_{t+\Delta t} - \hat{X}_{t+\Delta t})^T]$, describes the uncertainty related to $\hat{X}_{t+\Delta t}$; its expression is derived using the first order Taylor expansion of F , under the reasonable assumptions that X_t , ΔS and $\Delta\theta$ are uncorrelated and that their n^{th} order moments ($n > 2$) are negligible:

$$C_{t+\Delta t}^X = \frac{dF}{dX} C_t^X \frac{dF^T}{dX} + \frac{dF}{d\Delta S} C_t^{\Delta S} \frac{dF^T}{d\Delta S} + \frac{dF}{d\Delta\theta} C_t^{\Delta\theta} \frac{dF^T}{d\Delta\theta} + Q_t \quad (8)$$

where the partial derivatives of F are calculated at $(\hat{X}_t, \Delta S, \Delta\theta)$.

The first term expresses the uncertainty on $\hat{X}_{t+\Delta t}$ due to this on \hat{X}_t . The second and third terms are the variances of the measurement uncertainty on ΔS and $\Delta\theta$, which are inherent in the sensor (quantification,

uncertainty on R and B essentially). The covariances $C_t^{\Delta S}$ and $C_t^{\Delta \theta}$ were evaluated to be smaller than $6 \cdot 10^{-10} \text{ m}^2$ and $1 \cdot 10^{-8} \text{ rad}^2$ (respectively) during ΔT . So these terms are negligible in the expression of $C_{t+\Delta t}^X$.

The expression of Q_t is modeled by a diagonal matrix which terms are:

$$\begin{aligned} Q_{11} &= K_{ss} |\Delta S \cos \theta_t|, \\ Q_{22} &= K_{ss} |\Delta S \sin \theta_t|, \\ Q_{33} &= K_{s\theta} |\Delta S| + K_{\theta\theta} |\Delta \theta|, \end{aligned} \quad (9)$$

where K_{ss} is the odometry drifting coefficient along S , with respect to ΔS (m^2/m),

$K_{s\theta}$ is the odometry drifting coefficient along θ , with respect to ΔS , (rad^2/m),

$K_{\theta\theta}$ is the odometry drifting coefficient along θ , with respect to $\Delta \theta$, (rad^2/rad).

The coefficients K_{ss} , $K_{s\theta}$, $K_{\theta\theta}$ are determined by observation. The value of these coefficient depends on the robot and its environment, and on the confidence we need. These coefficient must be increased if the localization system has to be tolerant to unusual events. In our experiments they were fixed as follows:

$$K_{ss} = 0.001 \text{ m}^2/\text{m},$$

$$K_{s\theta} = 0.0003 \text{ rad}^2/\text{m},$$

$$K_{\theta\theta} = 0.001 \text{ rad}^2/\text{rad}.$$

These values are approximately two times the maximum drift we measured.

The coordinate system of the vehicle is defined figure 1. Notice that this coordinate system has its origin not at the back axle mid-point (X_A), but at the intersection of the camera pan rotation axis with the ground (X_R). The translation T , providing X_R as a function of X_A , is accompanied by an increase in the uncertainty, given by (11):

$$X_R = T(X_A), \quad (8)$$

$$C_t^{X_R} = \frac{dT}{dX}(X_A) C_t^{X_A} \left(\frac{dT}{dX}(X_A) \right)^T. \quad (9)$$

3. Visual observations

In order to proceed, we must make clear notation, coordinate systems and assumptions which we have used. We assume that the robot moves in the $O_s X_s Y_s$ plane of the (O_s, X_s, Y_s, Z_s) scene coordinate systems (figure 1). Its position is given at the time t by the state vector $X_t = (x_R, y_R, \theta_R)^T$.

The assembly of the camera has been mechanically adjusted so that the optical axis is kept horizontal. The angle φ defines the pan rotation. The rows of the CCD retina have been aligned with the horizontal defined by the ground plane. We assume that the CCD columns are vertical. A pin-hole model has been adopted for the camera. The origin O_c of the coordinate systems is at the optical center. The $O_c Z_c$ axis corresponds to the optical axis, and $O_c X_c$ (respectively $O_c Y_c$) to the rows (columns) of the CCD.

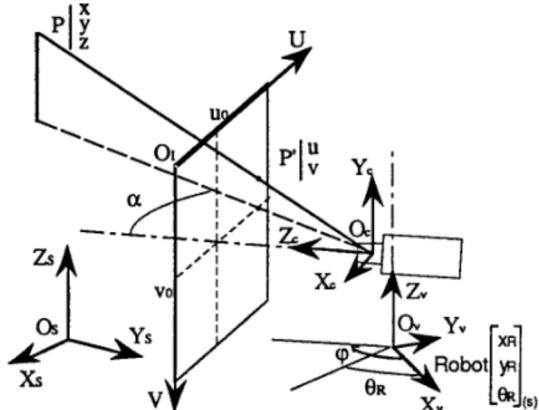


Figure 1 Definition of the four coordinate systems for the Scene, robot, camera and retina.

The coordinate system linked to the vehicle is defined as follows: $O_v Z_v$ is the pan rotation axis, O_v lying on the ground (plane $Z_s=0$). The direction $O_v X_v$ corresponds to the robot heading. The translation from camera to vehicle coordinates, $T = (T_x, T_y, T_z)^T$ is expressed in the scene coordinate system, the $O_v O_c$ vector when the pan angle φ equals zero. T is calculated during the visual system calibration.

The image plane coordinate system (O_i, U, V) is defined figure 1. Under these conditions, the equations giving the image coordinate $P'(u, v)$ of any scene point $P(x, y, z)$ projected onto the image are:

$$\frac{u-u_0}{\alpha_u} ((x-x_R)C + (y-y_R)S - T_x) - ((x-x_R)S - (y-y_R)C + T_y) = 0 \quad (12)$$

$$\frac{v-v_0}{\alpha_v} ((x-x_R)C + (y-y_R)S - T_x) - (T_z - z) = 0 \quad (13)$$

where $C = \cos(\varphi + \theta_R)$ and $S = \sin(\varphi + \theta_R)$, (u_0, v_0) are the image coordinate of the optical center projection, and α_u , α_v are the horizontal and vertical scale factors; u_0 , v_0 , α_u and α_v are calculated during calibration.

The identification of L landmarks at the time t=k gives the measurements vector $U_k = (u_1, \dots, u_n, v_1, \dots, v_p)^T$ ($n+p \geq L$). This observation is corrupted by the measurement noise Ψ_k , supposed to be Gaussian, zero-mean, with covariance matrix C_k^Ψ :

$$U_k = \hat{U}_k + \Psi_k, \text{ with } E[\Psi_k] = 0, \quad (14)$$

$$E[\Psi_k \Psi_k^T] = C_k^\Psi.$$

The measurement errors are due to visual feature extraction, un-perfected camera calibration, assumptions made about the coordinate systems, robot vibrations, uncertainty in the landmarks position, or other unpredictable factors. We suppose these noises to be uncorrelated, so that C_k^Ψ is diagonal. We assessed the measurement error standard deviation at the amount of 3 and 4 pixels for u and v respectively.

4. Updating the robot position

The purpose of one EKF iteration is to provide the estimated robot position at the time t=k, $\hat{X}_{k/k}^{(1)}$, with $\hat{X}_{k-1/k-1}$ and the measurements \hat{U}_k made at the time k. Moreover, the uncertainty related to this estimation is updated as well.

The EKF is based on a system equation, equation (7). This equation gives $\hat{X}_{k/k-1}$ at the instant a set of landmarks is identified. The covariance matrix $C_{k/k-1}^X$ associated to $\hat{X}_{k/k-1}$ is also provided, according to the process described in section 2.

The second part of the EKF, the measurement equation, is the system of ($n+p$) non linear equations described in (12) and (13), relating the robot position to the measurements vector; this system is written using the matrix form:

$$G(X_k, U_k) = 0. \quad (15)$$

This equation is linearized writing the first order Taylor expansion of G around $(\hat{X}_{k/k-1}, \hat{U}_k)$ (see [2]), which yields:

$$Y_k = H_k X_k + V_k, \quad (16)$$

where $Y_k = -G(\hat{X}_{k/k-1}, \hat{U}_k) + \frac{dG}{dX}(\hat{X}_{k/k-1}, \hat{U}_k) \hat{X}_{k/k-1}$,

$$H_k = \frac{dG}{dX}(\hat{X}_{k/k-1}, \hat{U}_k),$$

$$V_k = \frac{dG}{dU}(\hat{X}_{k/k-1}, \hat{U}_k) (U_k - \hat{U}_k),$$

$$E[V_k] = 0, \quad E[V_k V_k^T] = R_k \quad \text{and}$$

$$R_k = \frac{dG}{dU}(\hat{X}_{k/k-1}, \hat{U}_k) C_k^\Psi \left(\frac{dG}{dU}(\hat{X}_{k/k-1}, \hat{U}_k) \right)^T.$$

The updated estimate of X_k and its associated covariance are then classically given by:

$$\hat{X}_{k/k} = \hat{X}_{k/k-1} + K_k (Y_k - H_k \hat{X}_{k/k-1}) \quad (17)$$

$$C_{k/k}^X = (I - K_k H_k) C_{k/k-1}^X \quad (18)$$

where the Kalman gain, K_k is:

$$K_k = C_{k/k-1}^X H_k^T (H_k C_{k/k-1}^X H_k^T + R_k)^{-1}. \quad (19)$$

5. Experiments

Our experimental set-up uses a PC/AT 386 equipped with an image acquisition card. A four wheeled robot, powered by two electrical motors (one for each of the back wheels), is controlled by this PC via a RS 232 serial line. The data issued from the motor encoders arrives every two milliseconds through this connection. A CCD camera (756x581 pixels), equipped with a 12.5 mm focal lens, is mounted on the robot. A manual vernier permits to position it to a precision of a minute of arc. The experiment is performed in a $15 \times 15 \text{ m}^2$ room, in which 5 landmarks have been chosen. These landmarks correspond to support posts, a cable cache, and an electrical wire running down one of the walls. A grid pattern has been drawn on the floor, permitting us to verify the real robot position to a precision of 2 cm for X-Y and to the 0.3° for θ . The (x,y) coordinates of the 5 landmarks B_0, B_1, B_2, B_3, B_4 in the scene coordinate system are respectively (0,0), (7.2, 9.8), (7.2, 4.8), (-3.6, 4.5), (-3.6, 9.3) (in meters).

A landmark is detected as follows: knowing its position and $\hat{X}_{k/k-1}$ permits to point the camera in the approximate landmark direction. By supposing that X_k is Gaussian distributed with mean $\hat{X}_{k/k-1}$ and covariance $C_{k/k-1}^X$, an area can be defined in the image, where the landmark might be found with the probability P. Only this area is analyzed (edge extraction, followed by an Hough transform), which authorizes us rejecting aberrant measurements in advance, and furthermore limiting the

⁽¹⁾ Should be read "estimate of X made at the time k, using the information available up to time k"

processing. The Hough space analysis relies on the knowledge of the landmark form and position and on the robot location: a template of the "landmark Hough transform" is built, and we use a normalized correlation to measure the similarity between the Hough space and this template. We did not mention in the results the 5% to 10% cases where a landmark was not identified, this to test exclusively the estimator performances.

In a first set of experiments, the influence of the number of measurements on the localization precision was illustrated. The robot was placed at the position $X_k = [0.0m, 10.0m, -90.0^\circ]^T$. For each of the corrections listed in the table figure 2, we initialized the positioning system with

$$\hat{X}_{k/k-1} = \begin{bmatrix} -0.15m \\ 10.00m \\ -91.6^\circ \end{bmatrix}$$

and

$$C_{k/k-1}^X = \begin{bmatrix} (0.2m)^2 & 0 & 0 \\ 0 & (0.2m)^2 & 0 \\ 0 & 0 & (3.0^\circ)^2 \end{bmatrix}.$$

A very accurate precision can be obtained (better than 2 cm for XY, and 0.3° for θ , with 3 to 10 m remote landmarks spread around the vehicle) as long as a sufficient number of observations are made. The precision limits are due to the unavoidable measurements bias (camera calibration in particular). Finally, we verified that the filter convergence is nearly immediate, the updated position being the same (to the cm and to the 0.2°) after further iterations of the filtering.

In a second set of experiments, we defined a standard route composed of four stage: first a 2 m translation, then a 45° rotation, followed by a 3m translation, and finally a -90° rotation.

This route was executed several times, from the same initial position $X=(3.5, 11.0, -90.0)$ and with the initial uncertainty $(\sigma_x, \sigma_y, \sigma_\theta)=(0.05, 0.05, 0.5)$. During each execution we reduced the number of landmarks used. Results are given figure 3. The first row contains the standard deviations along x,y and θ given by our error model for the odometry, if the robot position and its related covariance is not corrected.

When a single landmark is detected, the standard deviations continues to grows. With two or more landmarks, the the standard deviations converge. We have here a rough indication about the required density of observations. Nevertheless, note that a good localization

was obtained using a single, but not always the same, landmark (experiment 7). Updating the robot position after each landmark detection, while frequently changing the located landmark, seems to be a good displacement strategy. Further experiments are however required to evaluate the interest of this approach.

The Gaussian hypothesis we have used is obviously only an approximation. None the less, we have made the χ^2 test described in [1] for the set of experiments presented here. The results show that the initial covariance we gave to the system is too large, but approaches the theoretical true value after several EKF iterations. We note that under estimating the covariance would lead to a loss of the robot's position. An over-estimation of the position covariance is necessary for stability of the technique.

6. Conclusion

We have presented a mobile robot localization system based on the use of odometry and the visual detection of common objects of the environment. The realistic model for the odometry and its associated uncertainty is the basis for a correct running of our system; the landmark detection as well as the updated position calculation rely on odometric estimates. A high adaptability, according to the number and the kind of landmarks required for the positioning, is provided by the combination of vision with an Kalman filter predict and update equations. Experimental results obtained with a first type of landmarks encourage us to extend the method to other objects of the environment. Future work concentrate on problems of generality and robustness of the vision techniques.

References

- [1] Y. Bar-Shalom, T. E. Fortmann, "Tracking and Data Association", New York: Academic, 1988.
- [2] N. Ayache, O. D. Faugeras, "Environment of a Mobile Robot", IEEE Trans. on Robotics and Automation, vol.5, n°6, pp 804-819, Dec. 1989.
- [3] J. L. Crowley, "World Modeling and Position Estimation for a Mobile Robot Using Ultrasonic Ranging", in Proc. IEEE Int. Conference on Robotics & Automation, pp 674-681, May 1989.
- [4] J. L. Crowley, "Control of Translation and Rotation in a Robot Vehicle", in Proc. IEEE Conference on Robotics & Automation, May 1989.

- [5] J.J. Leonard, H.F. Durant-Whyte, "Mobile Robot Localization by Tracking Geometric Beacons", IEEE Trans. on Robotics and Automation, Vol. 7, No. 3, pp 376-382, June 1991.
- [6] B. C. Bloom, "Use of Landmarks for Mobile Robot Navigation", SPIE vol. 579, Intelligent Robot & Computer Vision, pp 351-355, 1985.
- [7] S. Y. Chen, W. H. Tsai, "Determination of Robot Location by Common Shapes", IEEE Trans. on Robotics and Automation, Vol. 7, No. 2, pp , Feb. 1991.
- [8] M. R. Kabuka, E. Arenas, "Position Verification of a Mobile Robot Using Standard Pattern", IEEE Journal on Robotics & Automation, vol RA-3, n°6, pp 505-516, Dec. 1987.
- [9] Kriegman, D. J., E. Triendl, and T. O. Binford, "Stereo Vision and Navigation in Buildings for Mobile Robots", IEEE Transactions on Robotics and Automation, Vol 5, No 6, pp 792-803, Dec. 1989.
- [10] E. Krotkov, "Mobile Robot Localization Using a Single Image", in Proc. IEEE Int. Conference on Robotics and Automation, pp 978-983, 1989.
- [11] K. Sugihara, "Some Location Problems for Robot Navigation Using Single Camera", Computer Vision, Graphics, and Image Processing 42, pp 112-129, 1988.

Landmarks used	$\hat{x}_{k/k}$ [m]	$\hat{y}_{k/k}$ [m]	$\hat{\theta}_{k/k}$ [°]	$\sigma_{x k/k}$ [m]	$\sigma_{y k/k}$ [m]	$\sigma_{\theta k/k}$ [°]
B_0	-0.16	10.0	-91.0	0.19	0.20	1.10
B_0, B_1	-0.09	10.11	-90.7	0.15	0.12	0.93
B_0, B_1, B_2	-0.05	10.06	-90.5	0.13	0.09	0.82
B_0, B_1, B_2, B_3	-0.02	10.04	-90.3	0.09	0.06	0.54
B_0, B_1, B_2, B_3, B_4	0.00	10.02	-90.2	0.06	0.02	0.29

Figure 2 Comparison of the updated position obtained after 1, 2, 3, 4, and 5 pan angle measurements.

	stage 1		stage 2		stage 3		stage 4	
	Δx [m]	σ_x [m]						
1 Drifting of the odometry	0.05		0.05		0.15		0.12	
	0.07		0.07		0.15		0.15	
	1.71		2.50		3.30		4.10	
2 B_1 identified at each stage	-0.02	0.05	0.03	0.04	0.03	0.06	0.05	0.05
	0.01	0.06	0.00	0.06	0.02	0.09	0.03	0.09
	-0.3	0.79	-1.1	0.79	-0.2	1.62	-0.4	1.36
3 B_4 identified at each stage	-0.03	0.05	0.03	0.05	0.04	0.07	0.04	0.07
	0.01	0.06	-0.01	0.05	0.06	0.07	0.06	0.07
	0.0	0.53	0.1	0.55	-0.3	0.66	0.3	0.63
4 B_2 identified at each stage	0.03	0.05	0.03	0.05	0.03	0.05	0.04	0.05
	0.02	0.04	0.02	0.04	0.07	0.03	0.08	0.03
	-0.1	0.49	-0.1	0.51	-0.5	0.66	-1.3	0.70
5 B_1 and B_4 identified at each stage	0.01	0.02	0.00	0.01	-0.01	0.04	-0.01	0.03
	0.00	0.05	0.00	0.05	0.05	0.04	0.05	0.04
	0.2	0.23	0.0	0.22	-0.1	0.47	-0.3	0.37
6 B_0 and B_2 identified at each stage	0.03	0.04	0.01	0.03	0.02	0.04	0.02	0.03
	0.01	0.04	0.01	0.04	0.05	0.03	0.06	0.03
	0.1	0.34	0.0	0.35	-0.3	0.39	0.5	0.33
7 A single landmark at each stage, successively B_1 , B_0 , B_1 , B_0 .	0.01	0.04	0.02	0.04	0.01	0.05	0.02	0.05
	0.01	0.06	0.01	0.05	0.06	0.07	0.04	0.06
	0.8	0.79	0.5	0.40	-0.7	1.11	-0.1	0.36

Figure 3 Results of the experiments with the robot moving.